Kritische analyse: onderzoek naar gender-bevestigende zorg, spijt en psychisch welzijn

1. De veelgeciteerde "lage spijt"-cijfers

- Systematische review en meta-analyse (2021, *Plastic and Reconstructive Surgery*): 27 studies, 7.928 patiënten — gepoolde spijtprevalentie **1%** (95% BI <1%-2%).
- Vervolganalyse (2024): gepoolde schatting van **1,94%**.
- Amsterdam-cohortstudie: spijt (smal gedefinieerd als terugkeer voor hormonen om de transitie terug te draaien) bij **minder dan 1%.**
- Ouder onderzoek (Pfäfflin, 2.000 patiënten 1961-1991): spijt bij **minder dan 1%.**

Methodologische zwaktes van deze cijfers

- **Hoge uitval bij follow-up.** Mensen met spijt keren vaak niet terug naar de behandelende kliniek, en worden dus niet meegeteld. Huisartsen geven detransitie ook niet altijd door aan de oorspronkelijke kliniek.
- **Smalle definitie van "spijt".** Sommige studies tellen alleen "terugkeer voor hormonen om de transitie terug te draaien" als spijt — mensen die wel spijt voelen maar geen medische stap terug zetten, tellen niet mee.
- **Detransitie ≠ spijt.** Iemand kan stoppen om externe redenen (gebrek aan sociale steun, discriminatie, bijwerkingen) zonder de oorspronkelijke keuze als fout te ervaren, en omgekeerd.

De Bustos-meta-analyse (2021) — de bron achter bijna alle "1%"-cijfers, zelf zwaar bekritiseerd

Vrijwel elk "1%"-cijfer dat wereldwijd wordt aangehaald (in klinieken, voorlichtingsmateriaal, media) is terug te voeren op één meta-analyse (Bustos et al., 2021, *Plastic and Reconstructive Surgery Global Open*). Die studie is zelf, in een directe kritische briefwisseling in hetzelfde vakgebied, methodologisch zwaar onderuitgehaald:

- De auteurs erkennen zelf "matig-tot-hoog risico op vertekening (bias) bij sommige studies" — maar dat betreft in werkelijkheid **23 van de 27 geïncludeerde studies, oftewel 97% van alle 7.928 deelnemers.**
- Van de 27 studies scoorden slechts **5 studies (3% van alle deelnemers)** als methodologisch beter. Zelfs die 5 hadden nog **uitval bij follow-up van 28% tot meer dan 40%** — inclusief uitval door overlijden aan complicaties of suïcide, wat zelf mogelijk samenhangt met spijt en dus de uitkomst vertekent.
- Onafhankelijke reviewers (Brignardello-Petersen & Wiercioch, 2022) kwalificeerden de studie zelf als **"low-quality"**.
- Er werden concrete fouten gevonden: één studie werd verkeerd geïncludeerd (ging alleen over spijt over de gekozen operatietechniek, niet over de transitie zelf), en er waren data-extractiefouten die tot onjuiste conclusies leidden. Ook zou de zoekstrategie relevante, bekende studies hebben gemist.
- Belangrijk temporeel probleem: de meeste geïncludeerde studies dateren van jaren tot decennia geleden, terwijl de groep mensen die tegenwoordig chirurgie ondergaat qua samenstelling, leeftijd en selectiecriteria **substantieel anders** is — de criteria voor

toegang tot chirurgie zijn sindsdien flink versoepeld. Zelfs als het cijfer destijds klopte, zegt het dus weinig over de huidige praktijk.

- Een directe kritische reactie in de vakliteratuur concludeert expliciet: **"de bewering dat het spijtpercentage bekend is en laag is, is onjuist — het percentage is niet bekend."**

Conclusie: het "1%"-cijfer is geen breed gedragen, degelijk vaststaand feit — het is een veelgeciteerd getal met een wankele wetenschappelijke basis, gebouwd op grotendeels laagwaardige studies met hoge uitval. Dit is exact hetzelfde patroon als bij de NLGN4Y-studie (sectie over homoseksualiteit) en de Bränström & Pachankis-affaire verderop in dit document: een getal dat breed wordt herhaald voordat de onderliggende kwaliteit goed is gecontroleerd.

---

2. De Zweedse langetermijnstudie (Dhejne et al., 2011)

- Grootste, langste cohortstudie tot nu toe: 324 personen, follow-up 1973-2003 (tot 30 jaar), vergeleken met gematchte controlepersonen uit de algemene bevolking.

- **Hoofdbevinding (hele periode):** na chirurgie hadden deze personen circa **3x hogere algehele sterfte**, een sterk verhoogd suïciderisico, en meer psychiatrische opnames dan controlepersonen.
- **Cruciale nuance:** bij opsplitsing in twee periodes (1973-1988 vs. 1989-2003) **verdwenen** de verschillen in sterfte, suïcidepogingen en criminaliteit voor de latere groep grotendeels. Dit wijst op verbetering van zorg en/of afname van stigma over tijd, niet op een vaststaand, blijvend negatief effect van de behandeling zelf.
- **Wat de hoofdonderzoeker (Dhejne) zelf zei:** in een interview (2015) legde ze uit dat de latere groep (1989-2003) meer gebruikmaakte van geestelijke gezondheidszorg dan de controlegroep — wat ze "passend" noemde gezien aanhoudende discriminatie, niet als teken van mislukte behandeling. Ze verzette zich expliciet tegen het gebruik van haar studie als bewijs dat "transitie mensen suïcidaal maakt."
- **Fundamentele beperking van de studie:** er is **geen vergelijking gemaakt met mensen met gendersdysforie die geen behandeling kregen** — alleen met de algemene bevolking. Je kunt dus niet concluderen dat behandeling zélf de oorzaak is van verhoogde risico's; het kan evengoed liggen aan blijvende

minderheidsstress, discriminatie, of een al zwaardere psychiatrische voorgeschiedenis bij mensen die chirurgie zoeken.

- Concreet mechanisme (medicatie, controlefrequentie) achter de verbetering tussen de twee periodes is **niet gedocumenteerd** in de studie zelf — Dhejne noemt algemene verbetering van zorg en afname van stigma, geen specifiek protocol.

---

3. De Bränström & Pachankis-affaire (2019-2020) — een leerzaam voorbeeld

- Oorspronkelijke studie (2019, *American Journal of Psychiatry*, hele Zweedse populatie): concludeerde dat chirurgie samenhing met **minder** gebruik van geestelijke gezondheidszorg nadien — wereldwijd geciteerd als bewijs dat transitie helpt.
- Na kritische brieven van meerdere onderzoeksgroepen (Zweden, Noorwegen, VS, VK) over de statistische methodologie volgde een **officiële correctie** (2020): bij heranalyse, vergeleken met mensen met genderincongruentie die **geen** chirurgie

ondergingen, bleek er **geen aantoonbaar voordeel** van chirurgie.

- Bijkomende zwakte: minder dan 2% van de geopereerden had dit 10+ jaar geleden gedaan — feitelijk een kortetermijnstudie, ondanks dat het vaak als langetermijnbewijs werd gepresenteerd.
- Dit is een schoolvoorbeeld van hoe een gunstige, methodologisch zwakke claim eerst breed wordt overgenomen, en pas later gecorrigeerd — hetzelfde patroon als bij de eerder besproken replicatiecrisis.

---

4. De Cass Review (VK, 2024)

- Onafhankelijke, in opdracht van de NHS uitgevoerde review concludeerde dat de gepubliceerde onderzoeksbasis voor gender-bevestigende zorg bij jongeren van **lage methodologische kwaliteit** is — onvoldoende betrouwbaar voor stellige klinische aanbevelingen.
- Tegenreactie: WPATH, de Endocrine Society en een groep Yale-onderzoekers bekritiseerden de Review scherp, en stelden dat bewijskwaliteitsnormen inconsistent zijn toegepast. Dit debat loopt nog.

5. Waarom het "gouden standaard"-onderzoek niet bestaat — de ethische paradox

Het ideale onderzoek zou zijn: een grote groep mensen met gendersdysforie, uitgebreid gemeten vóór behandeling, **gerandomiseerd** verdeeld in een groep die behandeling krijgt en een groep die dat (nog) niet krijgt, met meting na 5, 10, 20 jaar in beide groepen. **Dit onderzoek bestaat niet en zal er waarschijnlijk ook nooit komen.**

De paradox uitgelegd

- Je kunt zo'n onderzoek alleen ethisch verantwoorden zolang je **niet zeker weet** of de behandeling werkt.
- Maar zodra klinieken, richtlijnen en beroepsverenigingen (WPATH, Endocrine Society) — wat al decennia het geval is — een **standpunt innemen** dat de behandeling waarschijnlijk helpt, wordt het als onethisch beschouwd om die behandeling alsnog aan een controlegroep te onthouden puur voor onderzoek.
- **Het gevolg:** het hardste bewijs komt er nooit, niet omdat niemand het wil onderzoeken, maar

omdat de sector al vroeg — vóórdat het sterke bewijs er was — een klinische positie innam. Die vroege positiekeuze sluit de deur naar het bewijs dat die positie zou moeten onderbouwen. Dit is een **self-fulfilling gap**: de onzekerheid blijft permanent bestaan, precies doordat men al besloten had dat de behandeling waarschijnlijk werkt.

Waarom dit "we behandelen al, dus verder onderzoek is onethisch"-argument zelf problematisch is

Dit verdient een aparte, kritische blik, want het wordt vaak als eindpunt van het debat gebruikt, terwijl het dat niet zou moeten zijn.

1. **De redenering die wordt gebruikt:** "We vermoeden dat de behandeling helpt tegen ernstig lijden → het zou onethisch zijn om dat vermoeden te testen met een onbehandelde controlegroep → dus behandelen we iedereen zonder harde toets → dus blijven we zeggen dat we weten dat het helpt."

2. **Waarom dit een self-fulfilling prophecy is.** Het "vermoeden" aan het begin van de redenering is nooit zelf hard getoetst. Het wordt binnengehaald als vaststaand feit, gebruikt om verder onderzoek te blokkeren, en vervolgens in

stand gehouden door het feit dát er geen tegenbewijs komt — terwijl dat gebrek aan tegenbewijs alleen bestaat omdát het onderzoek nooit is toegestaan. De cirkel sluit zichzelf: de aanname voorkomt de toets, en het ontbreken van de toets wordt gelezen als bevestiging van de aanname. Zo kun je nooit "bewijzen dat het niet werkt of zelfs schaadt" zolang iedereen al behandeld wordt vanuit de aanname dat het werkt — ongeacht of die aanname klopt.

3. **Waarom dát zelf onethisch kan zijn, niet alleen methodologisch zwak.**

    - Als de behandeling **wél** schadelijk of ineffectief blijkt te zijn voor een deel van de mensen (wat volledig mogelijk is zolang het niet getest is), dan worden op dit moment **grote groepen mensen onderworpen aan onomkeerbare, ingrijpende medische ingrepen op basis van een aanname die nooit hard getoetst is**. Dat keert het medisch-ethische basisprincipe om: normaal geldt "eerst goed bewijs, dan pas grootschalige toepassing"; hier is het "eerst grootschalig toepassen, en het bewijs vermijden omdat toepassing al gaande is."
    - Het beroep op "onethisch om te onthouden" beschermt in de praktijk vooral de positie van behandelaars en instellingen die al een

standpunt hebben ingenomen — niet per se de patiënt. De patiënt krijgt namelijk nooit de zekerheid die wél mogelijk was geweest met beter onderzoek vroeg in het proces (bijvoorbeeld wachtlijst-gerandomiseerde opzetten, die wél ethisch haalbaar zijn en soms ook zijn gebruikt, maar zelden met lange follow-up zijn voortgezet).

  - Zolang de onzekerheid blijft bestaan, blijft ook onbeantwoord of een deel van de behandelde mensen mogelijk **slechter af is geworden** dan zonder behandeling — en dat risico wordt afgedekt door een ethisch argument dat feitelijk vooral zichzelf beschermt tegen het achterhalen van die vraag, in plaats van de patiënt te beschermen.

4. **De vroege observationele data waarop de aanname destijds is gebaseerd, bleken bij latere kritische controle zwakker dan gepresenteerd** (zie de Bränström & Pachankis-correctie hierboven). Het fundament onder "we weten al dat het werkt" is dus zelf wankel — en juist dát wankele fundament wordt gebruikt om verder, steviger onderzoek af te wijzen als overbodig of onethisch.

5. **Vergelijking met andere geneeskunde relativeert dit gedeeltelijk, maar niet volledig.**

> Ook bij bariatrische chirurgie, sommige kankerbehandelingen en orgaantransplantatie is nooit een placebo-RCT gedaan, om dezelfde ethische reden. Maar bij die behandelingen was de fysiologische werking en de uitkomstmaat (overleving, tumorregressie) veel directer meetbaar. Bij gendersdysforie is de uitkomstmaat (psychisch welzijn, identiteitsbeleving) sterker beïnvloed door sociale context (stigma, discriminatie, sociale steun) — waardoor observationeel bewijs hier extra kwetsbaar is voor vertekening, en de self-fulfilling prophecy dus extra hardnekkig.

De eerlijkste conclusie

Het "we doen het al dus verder testen is onethisch"-verhaal is **geen bewijs dat de behandeling werkt** — het is een verklaring waarom we het *nooit hard zullen kunnen bewijzen of weerleggen* op de sterkste wetenschappelijke manier, en tegelijk een mechanisme dat zichzelf in stand houdt. Die twee dingen worden in het publieke debat vaak door elkaar gehaald: het ontbreken van tegenbewijs (omdat je het onderzoek niet mag doen) wordt gepresenteerd alsof het bevestigend bewijs is. Dat is dezelfde denkfout als eerder besproken bij "we hebben de oorzaak niet gevonden, dus het is normaal" — afwezigheid van bewijs in één richting is

geen bewijs vóór de andere richting. Het verschil hier is dat de lus zichzelf actief in stand houdt: de aanname blokkeert precies het onderzoek dat de aanname zou kunnen weerleggen.

Samenvattend oordeel

Onderdeel	Wat we denken te weten	Kernzwakte
Spijtcijfers (~1%)	Laag gerapporteerde spijt	Hoge uitval, smalle definitie van "spijt"
Zweedse studie (Dhejne 2011)	Verhoogde risico's, verbeterend over tijd	Geen onbehandelde controlegroep, mechanisme achter verbetering onbekend
Bränström & Pachankis (2019/2020)	Aanvankelijk "minder zorggebruik na chirurgie"	Na correctie: geen aantoonbaar voordeel; korte follow-up

Onderdeel	Wat we denken te weten	Kernzwakte
Cass Review (2024)	Bewijsbasis is methodologisch zwak	Zelf ook bekritiseerd door voorstanders van gender-bevestigende zorg
RCT / gouden standaard	Bestaat niet en komt er waarschijnlijk nooit	Ethisch geblokkeerd door vroege, zelf wankele aannames — een self-fulfilling gap

Kernconclusie: het eerlijkste antwoord op de vraag "maakt transitie mensen gelukkiger of ongelukkiger op de lange termijn, vergeleken met geen behandeling" is: **dat weten we op dit moment wetenschappelijk niet zeker.** Niet omdat er geen onderzoek is gedaan, maar omdat het sterkste type onderzoek (gerandomiseerd, met een onbehandelde controlegroep) ethisch is buitengesloten door een aanname die zelf nooit hard is getoetst — en de bestaande, zwakkere observationele data wijzen bij

kritische controle steeds weer minder eenduidig uit dan aanvankelijk gepresenteerd, in beide richtingen.